10 **Nakazawa M**, Kikawa, E, Chida Y, Tamai M. Asn244His mutation of the peripherin/RDS gene causing autosomal dominant cone-rod degeneration. Hum Mol Genet 1994;**3**:1195-6.

11 **Freund CL**, Gregory-Evans CY, Furukawa T, Papaioannou M, Looser J, Ploder L, Bellingham J, Ng D, Herbrick JAS, Duncan A, Scherer SW, Tsui LC, Loutradis-Anagnostou A, Jacobson SG, Cepko CL, Bhattacharya SS, McInnes RR. Cone-rod dystrophy due to mutations in a novel photoreceptor-specific homeobox gene (CRX) essential for maintenance of the photoreceptor. Cell 1997;**91**:543-53.

12 **Swain PK**, Chen SM, Wang QL, Affatigato LM, Coats CL, Brady KD, Fishman GA, Jacobson SG, Swaroop A, Stone E, Sieving PA, Zack DJ. Mutations in the cone-rod homeobox gene are associated with the cone-rod dystrophy photoreceptor degeneration. Neuron 1997;**19**:1329-36.

13 **Kelsell RE**, Gregory-Evans K, Payne AM, Perrault I, Kaplan J, Yang RB, Garbers DL, Bird AC, Moore AT, Hunt DM. Mutations in the retinal guanylate cyclase (RETGC-1) gene in dominant cone dystrophy. Hum Mol Genet 1998;**7**:1179-84.

14 **Downes SM**, Holder GE, Fitzke FW, Payne AM, Warren MJ, Bhattacharya SS, Bird AC. Autosomal dominant cone and cone-rod dystrophy with mutations in the guanylate cyclase activator 1A gene-encoding guanylate cyclase activating protein-1. Arch Ophthalmol 2001;**119**:96-105.

15 **Cremers FPM**, van de Pol DJR, van Driel M, den Hollander AI, van Haren FJJ, Knoers NVAM, Tijmes N, Bergen AAB, Rohrschneider K, Blankenagel A, Pinckers AJLC, Deutman AF, Hoyng CB. Autosomal recessive retinitis pigmentosa and cone-rod dystrophy caused by splice site mutations in the Stargardt's disease gene ABCR. Hum Mol Genet 1998;**7**:355-62.

16 **Wissinger B**, Gamer D, Jagle H, Giorda R, Marx T, Mayer S, Tippmann S, Broghammer M, Jurklies B, Rosenberg T, Jacobson SG, Sener EC, Tatlipinar S, Hoyng CB, Castellan C, Bitoun P, Andreasson S, Rudolph G, Kellner U, Lorenz B, Wolff G, Verellen-Dumoulin C, Schwartz M, Cremers FP, Apfelstedt-Sylla E, Zrenner E, Salati R, Sharpe LT, Kohl S. CNGA3 mutations in hereditary cone photoreceptor disorders. Am J Hum Genet 2001;**69**:722-37.

17 **Roepman R**, Bernoud-Hubac N, Schick DE, Maugeri A, Berger W, Ropers HH, Cremers FP, Ferreira PA. The retinitis pigmentosa GTPase regulator (RPGR) interacts with novel transport-like proteins in the outer segments of rod photoreceptors. Hum Mol Genet 2000;**9**:2095-105.

18 **Zhao Y**, Hong DH, Pawlyk B, Yue G, Adamian M, Grynberg M, Godzik A, Li T. The retinitis pigmentosa GTPase regulator (RPGR)-interacting protein: Subserving RPGR function and participating in disk morphogenesis. Proc Natl Acad Sci USA 2003;**100**:3965-70.

19 **Gerber S**, Perrault I, Hanein S, Barbet F, Ducroq D, Ghazi I, Martin-Coignard D, Leowski C, Homfray T, Dufier JL, Munnich A, Kaplan J, Rozet JM. Complete exon-intron structure of the RPGR-interacting protein (RPGRIP1) gene allows the identification of mutations underlying Leber congenital amaurosis. Eur J Hum Genet 2001;**9**:561-71.

20 **Dryja TP**, Adams SM, Grimsby JL, McGee TL, Hong DH, Li T, Andreasson S, Berson EL. Null RPGRIP1 alleles in patients with Leber congenital amaurosis. Am J Hum Genet 2001;**68**:1295-8.

21 **Underhill PA**, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ. Y chromosome sequence variation and the history of human populations. Nat Genet 2000;**26**:358-61.

22 **Cremers FP**, van den Hurk JA, den Hollander AI. Molecular genetics of Leber congenital amaurosis. Hum Mol Genet 2002;**11**:1169-76.

Geographical and ethnic variation of the 677C>T allele of 5,10 methylenetetrahydrofolate reductase (*MTHFR*): findings from over 7000 newborns from 16 areas world wide

B Wilcken, F Bamforth, Z Li, H Zhu, A Ritvanen, M Redlund, C Stoll, Y Alembik, B Dott, A E Czeizel, Z Gelman-Kohan, G Scarano, S Bianca, G Ettore, R Tenconi, S Bellato, I Scala, O M Mutchinick, M A López, H de Walle, R Hofstra, L Joutchenko, L Kavteladze, E Bermejo, M L Martínez-Frías, M Gallagher*, J D Erickson*, S E Vollset*, P Mastroiacovo*, G Andria*, L D Botto*

....

Since its biochemical characterisation in 1991[1] and its genetic identification in 1995,[2] 677C>T allele (T allele) of the 5,10 methylenetetrahydrofolate reductase (*MTHFR*) gene has been a focus of increasing interest from researchers world wide. The expanding spectrum of common conditions linked with the 677C>T allele now includes certain adverse birth outcomes (including birth defects), pregnancy complications, cancers, adult cardiovascular diseases, and psychiatric disorders.[3–8] Although several of these associations remain unconfirmed or controversial,[4] their scope is such that it becomes of interest to explore the geographical and ethnic distribution of the allele and associated genotypes.[9] Accurate information on such distribution can contribute to studies of gene-disease associations (by providing reference population data) and population genetics (by highlighting geographical and ethnic variations suggestive of evolutionary pressures),[10] as well as help to evaluate health impact (by allowing estimates of population attributable fraction).

Current population data, however, show gaps and even for some ethnic groups or large geographical areas (for example,

China) few data are available.[3] Our aim was to supplement the available data by collecting a large and diverse sample of newborns from different geographical areas and ethnic groups, and to examine international variations in the distribution of the 677C>T allele. We present findings relating to more than 7000 newborns from 16 areas around the world.

MATERIALS AND METHODS

The study was conducted under the auspices of the International Clearinghouse for Birth Defect Monitoring Systems (ICBDMS) and was coordinated through its head office, the International Center on Birth Defects (ICBD).

Sample selection

Participating programmes, in consultation with the coordinating group, identified a population sampling approach that would be simple yet minimise sampling bias with respect to the *MTHFR* genotype. We made an explicit attempt to sample systematically the newborn population. Details of each programme's approach are listed below, and further information is available on request.

Generally, programmes chose one of two approaches. The first approach used regional newborn screening programmes as the source of samples. Typically, such an approach used a

....

Authors are listed by alphabetical order of country with coordinating group and corresponding author last. *Coordinating group.

Key points

- Our objectives were to characterise the geographical and ethnic distribution of the 677C>T allele (T allele) of the *MTHFR* gene and its associated genotypes among newborns around the world, using newborn screening programs and birth hospitals. The participants were 7130 newborns of different ethnicities from 16 areas in Europe, Asia, the Americas, the Middle East, and Australia.

- The distribution of the allele showed marked ethnic and geographical variation The homozygous TT genotype was particularly common in northern China (20%), southern Italy (26%), and Mexico (32%). There was also some evidence for geographical gradients in Europe (north to south increase) and China (north to south decrease). The TT genotype frequency was low among newborns of African ancestry, intermediate among newborns of European origin, and high among newborns of American Hispanic ancestry. Areas at the extremes of the frequency distribution showed deviations from Hardy-Weinberg expectations (Helsinki, Finland, southern Italy, and southern China).

- This study, the largest to date, suggests the presence of selective pressures leading to marked geographical and ethnic variation in the frequency of the 677C>T allele. Geneticists can benefit from these reference data when examining links between the 677C>T allele and health outcomes in diverse populations.

geographically defined birth population. In Atlanta, for example, researchers visited the Georgia newborn screening programme on different days over several weeks and selected a 1 day collection of blood spots received by the laboratory from children whose mothers resided in one of five counties in Atlanta. Discussions with the director of the newborn screening programme indicated that day to day variability in the flow of specimens from birth hospitals to the state laboratory was negligible. The second approach relied on systematic sampling directly from birth hospitals that were part of an established network. In Spain, for example, staff collected specimens from 15 consecutive newborns at each hospital participating in the ECEMC monitoring programme, which includes birth hospitals from across Spain. Details for specific areas are summarised below.

Australia, New South Wales

Specimens were obtained from the New South Wales newborn screening programme, by selecting 100 consecutive newborn screening cards on each of five consecutive days, excluding repeat specimens, for a total of 500 specimens. All maternity units in the state of New South Wales send their specimens to the programme. Specimens consisted of blood remaining after routine newborn screening tests had been completed.

Canada, Alberta

Specimens were taken from consecutive newborns from the provincial newborn screening programme in Alberta. Specimens consisted of the remaining blood spots used by the newborn screening programme. The first 100 specimens of the month were collected each month for four months.

China, northern and southern

Umbilical cord blood samples were collected from newborns from major hospitals in 12 cities in China from March to November 1998. One hundred consecutive samples were requested from each hospital. The hospitals were in cities from southern China (Wuhan, Nanjing, Guangzhou, and Chengdu) and northern China (Yanbian, Urumchi, Changchun, Jinan, Xi'an, Shenyang, Beijing, and Jilin). For homogeneity, only newborns of Han ethnicity were included in the study.

Finland, Helsinki

Specimens originated from newborns in the major maternity hospital in Finland, at the Helsinki University Hospital. Sampling was restricted to babies whose parents were both Finns. Specimens consisted of the remainder of umbilical blood specimens for hypothyroidism screening. The latter are collected for every newborn in Finland.

France, Strasbourg

Specimens were taken from consecutive newborns from newborn screening centres in Département du Bas-Rhin, whose births are also covered by the Strasbourg Birth Defects Registry. Specimens consisted of the remainder of blood spots used by the newborn screening programme.

Hungary

Specimens were taken from consecutive newborns. Specimens were collected from the remainder of the blood spots from the two newborn screening centres that operate in Hungary. For twin pairs, only one of the pair, selected at random, was included.

Israel, Tel Aviv

Specimens were taken from consecutive newborns from one major university hospital in Tel Aviv and consisted of blood spots.

Italy, Campania

Specimens were taken from consecutive newborns at three hospitals in Campania (two in Avellino, one in Benevento). Specimens consisted of the remainder of blood spots used by the newborn screening programme.

Italy, Sicily

Specimens were taken from consecutive newborns from the newborn screening programme in south east Sicily. Specimens consisted of the remainder of blood spots used by the newborn screening programme.

Italy, Veneto

Specimens were taken from consecutive newborns at one hospital outside the town of Vicenza. The hospital was chosen because it is an area hospital with 1200 births per year that has good obstetric care but does not select for high risk pregnancies. Specimens consisted of the remainder of blood spots used by the newborn screening programme.

Mexico

Specimens were randomly selected from blood spots from newborns born in hospitals that are part of the RYVEMCE birth defect monitoring network in Mexico. Samples were obtained from the remainder of the blood spot specimens collected for hypothyroidism screening. Selection was stratified to include equal numbers of males and females in the final sample.

The Netherlands, northern region

Specimens were randomly chosen from newborns whose mothers resided in the northern Netherlands. Specimens consisted of the remainder of blood spots used by the newborn screening programme.

Spain

Specimens were taken from consecutive newborns in 67 hospitals of the National Health Service throughout Spain. Essentially all babies in Spain are born in such hospitals. These hospitals are part of ECEMC (Spanish Collaborative Study of

Table 1 Allele and genotype frequency of the 677C>T allele of the *MTHFR* gene by area and ethnic origin

Area	Sample No	Genotype (No)* CC	CT	TT	Genotype (%) CC (%)	CT (%)	TT (%)	95% CI	Allele frequency T (%)
Europe									
Italy, Sicily	468	138	236	94	29	50	**20.1**	16.7 to 24.0	**45.3**
Italy, Campania	500	172	196	132	34	39	**26.4**	22.7 to 30.4	**46.0**
Italy, Veneto									
Italian whites	385	128	198	59	33	51	**15.3**	12.1 to 19.3	**41.0**
Others	47	27	15	5	57	32	**10.6**	4.6 to 22.6	**26.6**
Spain, multicentre									
Spanish whites	601	265	265	71	44	44	**11.8**	9.5 to 14.6	**33.9**
Others	51	24	20	7	47	39	**13.7**	6.8 to 25.7	**33.3**
France, Strasbourg	178	72	85	21	40	48	**11.8**	7.8 to 17.4	**35.7**
The Netherlands	188	97	79	12	52	42	**6.4**	3.7 to 10.9	**27.4**
Finland, Helsinki	545	293	230	22	54	42	**4.0**	2.7 to 6.0	**25.1**
Hungary	378	165	171	42	44	45	**11.1**	8.3 to 14.7	**33.7**
Russia	587	312	234	41	53	40	**7.0**	5.2 to 9.3	**26.9**
Middle East									
Israel	210	120	72	18	57	34	**8.6**	5.5 to 13.1	**25.7**
China									
North, Han ethnicity	643	201	315	127	31	49	**19.8**	16.9 to 23.0	**44.2**
South, Han ethnicity	430	167	228	35	39	53	**8.1**	5.9 to 11.1	**34.7**
Oceania									
Australia									
Whites	288	146	119	23	51	41	**8.0**	5.4 to 11.7	**28.6**
Others	75	40	27	8	53	36	**10.7**	5.5 to 19.7	**28.7**
Americas									
Mexico	500	91	248	161	18	50	**32.2**	28.3 to 36.4	**57.0**
USA, Atlanta									
Whites	300	142	126	32	47	42	**10.7**	7.7 to 14.7	**31.7**
Blacks	298	231	59	8	78	20	**2.7**	1.4 to 5.2	**12.6**
Hispanics	62	22	29	11	35	47	**17.7**	10.2 to 29.0	**41.1**
Asian	26	16	9	1	62	35	**3.8**	0.7 to 18.9	**21.2**
Other, unknown	100	63	32	5	63	32	**5.0**	2.2 to 11.2	**21.0**
Canada, Alberta									
Whites	240	136	90	14	57	38	**5.8**	3.5 to 9.6	**24.6**
Others	30	17	13	0	57	43	**0.0**	0.0 to 11.4	**21.7**
Total	**7130**								

*TT, two 677C>T alleles; CT, one 677C>T allele; CC, no 677C>T allele.

Congenital Malformations), which monitors one quarter of all births in Spain. Each hospital contributed specimens for 15 consecutive newborn infants during three selected months. Specimens were collected at the same time as the blood spots for the newborn screening programme.

Russia
Specimens were selected from the neonatal screening programme that collects and banks specimens from 54 maternity hospitals in the Moscow area. All selected babies were apparently free from congenital anomalies.

USA, Atlanta (Georgia)
Specimens were chosen from newborns whose mothers resided in one of the five counties in metropolitan Atlanta, as ascertained from information on the newborn screening card. Staff visited the newborn screening programme four times over two months. At each visit researchers selected at random the specimens collected during one day and collected specimens from the blood spots left over from newborn screening.

Sample determination and data collection
We determined that a sample size of approximately 400 to 500 specimens per area would provide a reasonably precise estimate (plus or minus 3%) for a genotype with 10% frequency. Such a frequency is within the range reported for the homozygous 677C>T (TT) genotype in many European countries and among North Americans of European descent, and is intermediate between the lower frequencies reported in some populations of African descent, and the higher frequencies reported in specimens from Mexico, Italy, and Hispanics in the USA.[3] Although 400 to 500 was the targeted number of specimens per area, smaller sample sizes were accepted, recognising that such samples would provide less precise estimates. For each sample, researchers collected information on sex and race/ethnicity. Ethnicity was determined from the blood spot card (for example, in Atlanta), from maternal interview (for example, Italy, Veneto, Spain), or from the last name or birth place of the parent (for example, Australia). Not all programmes collected all variables.

Human subject protection
Local review boards approved the study. In most cases specimens were anonymised before testing. In all cases, personal identifiers were removed before data were provided to ICBD for epidemiological analysis.

Laboratory testing
Genomic DNA was isolated from blood spots collected on filter paper. The presence of the C>T change within the *MTHFR* gene creates a *Hin*fI restriction site that can be detected by restriction enzyme digestion followed by electrophoresis. Amplification of the *MTHFR* gene by the polymerase chain reaction and detection of the T allele was performed using protocols based on the method of Frosst *et al*.[2]

Five programmes (USA, China, Israel, Mexico, The Netherlands) tested their own specimens. All other specimens were tested at a single laboratory (Naples, Italy). The laboratories participated in proficiency testing to ensure inter-laboratory
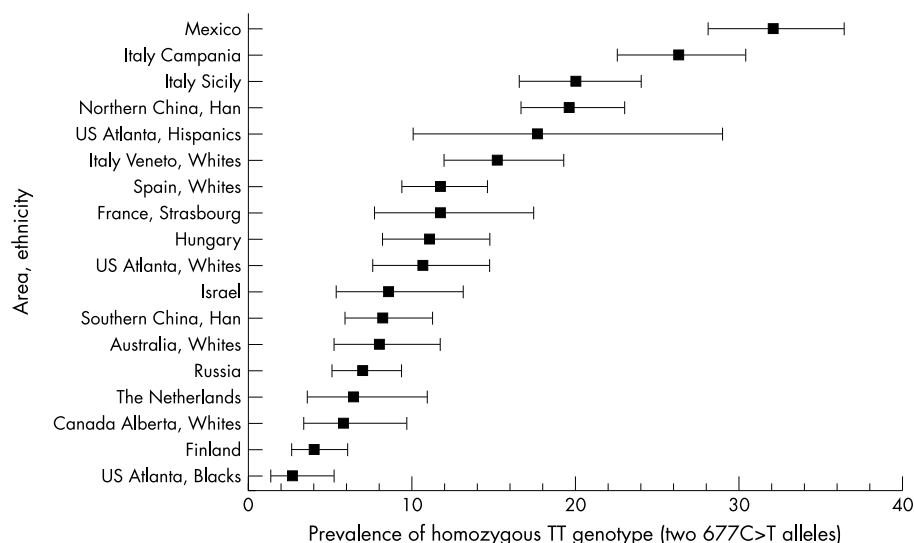
Figure 1 Prevalence of homozygous TT genotype (two 677C>T alleles) among newborns by area and ethnic background, ICBDMS 2003.

consistency. The proficiency testing consisted of the preparation of punches from 12 blood spots in one laboratory (CDC) with a mix of genotypes (CC, CT, and TT). These genotypes were confirmed by sequencing. Punches from these blood spots were then sent to the other laboratories, which performed DNA extraction followed by genotype assay. The laboratories were blinded to the genotype of these specimens as well as to the relative proportion of the genotypes. Results from each laboratory were sent back to CDC for evaluation. With the exception of one sample, which could not be amplified by four of the six laboratories, results showed complete agreement across laboratories.

Statistical analysis
We computed the confidence intervals for proportions using the Wilson score method without continuity correction.[11][12] Deviation from Hardy-Weinberg equilibrium was tested by chi-square analysis. In addition to allele frequencies, we present genotype frequencies for the homozygous TT genotype (two 677C>T alleles), the heterozygous CT genotype (one 677C>T allele), and the homozygous CC genotype (no 677C>T alleles).

RESULTS
We present data on 7130 newborns from 16 areas in the Americas, Europe, Russia, China, and Australia. Amplification rates for blood spots by geographical source were the following: Italy, Sicily 89%, Italy, Campania 76%, Italy, Veneto 73%, Spain 69%, France, Strasbourg 78%, Finland, Helsinki 95%, Hungary 95%; Russia, Moscow 95%, Australia, New South Wales 79%, Canada, Alberta 77%, USA, Atlanta 96%. By comparison, a large Irish population based study using newborn blood spots successfully genotyped 85% of collected samples.[13]

Prevalence by geographical area and ethnicity
The allele and genotype distribution, by area and ethnicity, is presented in table 1. The prevalence of the homozygous TT genotype (two 677C>T alleles) is visually summarised in fig 1 for those groups with at least 50 subjects tested.

The distribution of the 677C>T allele showed regional and ethnic variations. For example, the prevalence of the homozygous TT genotype was 10-12% in several areas in Europe (for example, Spain, France, and Hungary). However, the prevalence appeared to be lower (4% and 6%, respectively) in Finland, Helsinki and the northern Netherlands, whereas in

some areas in southern Europe it was much higher (26% and 20% in Campania and Sicily, respectively). In the Americas, the frequency of the homozygous TT genotype was higher in Mexico (32%), intermediate in Atlanta (11% among whites), and somewhat lower in Alberta (6%). In Australia, TT prevalence was 7.5% among whites.

Genotype varied by ethnicity as well as by geographical location. For example, TT homozygosity was more common among newborns from Mexico or those born in Atlanta of Hispanic origin, intermediate among newborns of European ancestry (for example, in Europe and North America), and lower among newborns of African ancestry (for example, in Atlanta and Veneto, Italy). However, a range of genotype frequencies was evident even within broad ethnic groups. For example, TT homozygosity among whites ranged from as low as 6% in Alberta (Canada), to 7.5% in New South Wales (Australia), to 11% in Atlanta (USA), to the high values already noted for Italy. For other ethnic and racial groups, such estimates are more unstable because of the smaller number of specimens, but it is worth noting an apparently low frequency of TT homozygosity among newborns of Asian origin from Australia and Atlanta.

Hardy-Weinberg equilibrium
The observed distribution of the three genotypes (CC, CT, TT) in most areas was similar to that expected under Hardy-Weinberg equilibrium (table 2). This was true also among males and females separately (data not shown). We found a relative excess of TT homozygotes in Campania, Italy, and an excess of heterozygotes in Helsinki, Finland and southern China (p<0.05).

DISCUSSION
We documented distinctive geographical and racial/ethnic variation in the prevalence of the 677C>T allele of the *MTHFR* gene among a large international sample of newborns. The several fold variation in the prevalence of the TT homozygous genotype across the study areas (fig 1) was also consistent, in some areas, with the presence of geographical gradients. In Europe, for example, the prevalence of the TT genotype increased in a roughly southerly direction, from low values in the north (4-7% in Finland, Helsinki, northern Netherlands, and Russia), to intermediate values (8-10%) in France and Hungary, to higher values in southern Europe (12-15% in Spain and northern Italy), peaking in southern Italy (20-26% in Campania and Sicily). In North America, the frequency of

Table 2 Distribution of *MTHFR* genotypes, Hardy-Weinberg expectations, and p value

Area	Sample No	Genotype (No)			Observed/expected			p value, χ^2
		CC	CT	TT	CC	CT	TT	
Europe								
Italy, Sicily	468	138	236	94	1.0	1.0	1.0	
Italy, Campania	500	172	196	132	1.2	0.8	1.2	p <0.05
Italy, Veneto								
Italian whites	385	128	198	59	1.0	1.1	0.9	
Spain, multicentre								
Spanish whites	601	265	265	71	1.0	1.0	1.0	
France, Strasbourg	178	72	85	21	1.0	1.0	0.9	
The Netherlands	188	97	79	12	1.0	1.1	0.9	
Finland, Helsinki	545	293	230	22	1.0	1.1	0.6	p <0.05
Hungary	378	165	171	42	1.0	1.0	1.0	
Russia	587	312	234	41	1.0	1.0	1.0	
Middle East								
Israel	210	120	72	18	1.0	0.9	1.3	
China								
North, Han	643	201	315	127	1.0	1.0	1.0	
South, Han	430	167	228	35	0.9	1.2	0.7	p <0.05
Oceania								
Australia								
Whites	288	146	119	23	1.0	1.0	1.0	
Americas								
Mexico	500	91	248	161	1.0	1.0	1.0	
USA, Atlanta								
Whites	300	142	126	32	1.0	1.0	1.1	
Blacks	298	231	59	8	1.0	0.9	1.7	
Hispanics	62	22	29	11	1.0	1.0	1.0	
Asian	26	16	9	1	1.0	1.0	0.9	
Other, unknown	100	63	32	5	1.0	1.0	1.1	
Canada, Alberta								
Whites	240	136	90	14	1.0	1.0	1.0	

TT homozygotes increased from western Canada (Alberta) to south eastern United States (Atlanta) and peaked in Mexico.

Ethnic variation was apparent among and within geographical areas. In metropolitan Atlanta, for example, TT homozygosity was common among newborns of Hispanic origin (15%), intermediate among those of European origin (11%), and low among African-American newborns (3%). These data are consistent with the high prevalence of TT homozygosity among newborns from Mexico in this study and with published data from the population based sample of babies of Mexican ancestry from California.[14] The low prevalence among US blacks is similar to that reported in pooled estimates of five studies on US blacks and three studies from sub-Saharan Africa[3] as well as in later studies from South Africa and Zimbabwe.[15 16] The intermediate prevalence among whites in Atlanta is consistent with similar rates observed in several European areas in this and several other studies.[3] However, more detailed comparisons are difficult because of the misclassification and imprecision of such ethnic labels.

Of note is the finding in Australia of a lower prevalence of the TT genotype among whites (7.5%) compared to previous reports.[17] Also, we noted a relatively low prevalence of the TT genotype (5.8%) among the random sample of white newborns in Alberta (Canada), compared to the frequency (11%) reported in a previous study from Quebec (Canada).[18] The latter study differed from ours in that newborns were enrolled from a single university hospital in Montreal and were selected, by design, so that their birth weights were at or above the 10th centile.[18]

The high frequency of TT homozygotes observed in this study among newborns from Mexico, northern China, and southern Italy was notable. These findings confirm and extend those previously reported from Mexico[19] and southern Italy.[20] Why such high rates of TT homozygosity occur in these regions is unclear, given the apparently limited ethnic, genetic,

or environmental commonalities among such areas. Researchers have suggested the possibility of heterozygote advantage with respect to the risk for neural tube defects.[21] However, such a hypothesis remains unconfirmed. Nevertheless, further exploration of gene-gene and gene-environment interaction might help to identify the evolutionary pressures favouring a high prevalence of this gene variant in certain areas and ethnic groups.

The impact of such geographical and ethnic variation on the distribution of disease in the population is unclear. For example, one would predict high rates of neural tube defects, whose risk appears to be increased nearly two-fold in the presence of 677C>T homozygosity[3] in those geographical areas or ethnic groups with a high frequency of this genotype. The evidence supporting such relations is mixed. For example, the data are consistent for Mexico and northern China, which not only have a very high frequency of the TT genotype but also high rates of neural tube defect.[22 23] Furthermore, within China, rates of neural tube defect are higher in the north (where the TT homozygous genotype is more common) than in the south.[23] In the United States, the rates of neural tube defects historically have been higher among Hispanics, intermediate among non-Hispanic whites, and lower among African-Americans, a trend that follows the relative frequency of the TT homozygous genotype.

There are, however, notable exceptions. In southern Italy, for example, the TT genotype is common, but the rate of neural tube defects is not particularly high.[22] Nevertheless, such exceptions are not entirely unexpected, because environmental and nutritional factors are likely to modulate considerably the genetic risk for neural tube defects. In fact, these exceptions might prove particularly valuable when investigating the aetiological heterogeneity and the role of interactions in the occurrence of neural tube defects.

Similar analyses are possible with respect to other outcomes. For example, recent meta-analyses showed associations of the TT genotype with ischaemic heart disease, deep

venous thrombosis, and perhaps stroke.[24] Like neural tube defects, these health outcomes are subject to interacting risk factors and therefore the relation between genotype and outcome at a group level is likely to be complex. Nevertheless, researchers seeking to understand such relations might find data such as these on the geographical and ethnic variation of the 677C>T allele helpful.

On a population level, the genotype distribution associated with the T allele was generally consistent with Hardy-Weinberg expectations. However, a few significant deviations did occur, mostly at the ends of the frequency spectrum. An excess of TT homozygotes was observed in southern Italy (Campania), where the allele was common, whereas the reverse was observed in Finland (Helsinki) where the allele was uncommon. Though these two deviations from Hardy-Weinberg expectations could be the result of chance and multiple statistical testing, they might also suggest the presence of local selective pressures.

In interpreting the findings of this study, one should consider its strengths and limitations. Although we attempted to draw unbiased, systematic samples of newborns from defined populations, sampling strategies varied across areas, and one cannot be certain that the efforts were always entirely successful. We provide details on sampling procedures as guidance to readers who wish to use part or all of these data. Dealing with race and ethnicity was also a difficult but inescapable challenge. Classifications based on self report and particularly on the birth place of the parent or last name are unsatisfactory to varying degrees. Thus, we present our data (table 1) either stratified in two groups (the main ethnic group and all other groups combined), or present data only for the major ethnic group (for example, Han Chinese). While this approach does not solve the difficulties entirely, it decreases the misclassification inherent in defining the many smaller ethnic groups that coexist in many areas. Other limitations of this study include the lack of coverage from many areas of the world, including most of Africa, the Middle East, Latin America, and the Indian subcontinent.[25]

Another challenge of this study was addressing measurement error in genotyping. One might speculate, for example, that deviations in Hardy-Weinberg equilibrium may be the result of genotyping errors. However, inter-laboratory consistency and quality control measures showed remarkable agreement among laboratories. In addition, the same laboratory that assayed samples from areas showing deviations from Hardy-Weinberg equilibrium also assayed the samples from many areas not showing such deviations, suggesting no systematic laboratory error.

A strength of the study was the ability to assemble systematically relatively large samples from newborns using explicit sampling protocols. Measures were also taken to ensure the reliability and comparability of genotypic data across laboratories, including quality control protocols that involved blind retesting of results and exchange of specimens.

Data from studies such as these can serve several purposes. Geneticists could find them useful when evaluating the distribution of genetic variation in human populations and its role in genetic susceptibility to disease. For example, population data might help geneticists reassess controversial associations such as that between *MTHFR* genotypes and risk for Down syndrome,[26–29] for which the evidence favouring the association was largely derived from comparisons with convenient samples of controls. As discussed previously, these genetic data can help to interpret prevalence gradients of disease, such as the well known geographical gradients of neural tube defect occurrence. Similarly, huge amounts of data on other outcomes, such as other birth defects, pregnancy complications, certain cancers, adult cardiovascular disease, and certain psychiatric disorders,[3–8] could be called upon to interpret the prevalence gradients noted in this and other studies. Our data are offered as a contribution to such investigation.

Population data on the 677C>T variant might also help population and public health geneticists assess the potential impact of preventive measures based on environmental modifications. For example, some adverse biochemical effects of the thermolabile enzyme coded by the T allele, such as the increase in total plasma homocysteine, appear to be reversible by increasing the consumption of the B vitamin folic acid.[30] If the effect of folic acid varies by genotype, then the overall impact in the population of fortification or supplementation programmes might vary predictably once the genotype distribution is known.

Finally, a practical outcome of this collaborative study was to show the feasibility of conducting such genetic surveys using existing networks of hospitals, birth defect registries, and research institutions. Other research groups have carefully selected and examined large and representative samples of newborns from single states or countries (for example, California[14] and Ireland[13]) and generated genotype prevalence data. We tried to expand such efforts to an international scale, and suggest that, with appropriate planning, such international networks can use their access and experience in community based studies to provide core data on the population distribution of common gene variants. These data in turn can serve as the foundation for studies of genetic variation and its role in increasing or decreasing disease risk.

. . . .

Authors' affiliations

B Wilcken, Children's Hospital at Westmead, Sydney, Australia
F Bamforth, University of Alberta, Edmonton, Canada
Z Li, H Zhu, National Centre for Maternal and Infant Health, Peking University, Beijing, China
A Ritvanen, The National Research and Development Centre for Welfare and Health, STAKES, Finland
M Redlund, Department of Paediatrics, Helsinki University Central Hospital, Finland
C Stoll, Y Alembik, B Dott, University of Strasbourg, Strasbourg, France
A E Czeizel, Foundation for the Community Control of Hereditary Diseases, Budapest, Hungary
Z Gelman-Kohan, Kaplan Medical Centre, Rehovot, Israel
G Scarano, Division of Medical Genetics, AO "G Rummo", Benevento, and Birth Defects Registry of Campania, Italy

S Bianca, G Ettore, Sicily Congenital Malformation Registry, Sicily, Italy
R Tenconi, Department of Paediatrics, University of Padova, Veneto, Italy
S Bellato, Paediatrics Unit, Arzignano Hospital, Vicenza, Veneto, Italy
I Scala, G Andria, University Federico II, Naples, Italy
O M Mutchinick, M A López, Departamento de Genética, RYVEMCE, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, México, DF
H de Walle, R Hofstra , University of Groningen, The Netherlands
L Joutchenki, L Kavteladze, Moscow Birth Defects Registry, Moscow, Russia
E Bermejo, ECEMC and Instituto de Salud Carlos III, Madrid, Spain
M L Martínez-Frías, ECEMC, Instituto de Salud Carlos III, and Universidad Complutense, Madrid, Spain
M Gallagher, National Center on Environmental Health, Centers for Disease Control and Prevention, Atlanta, Georgia, USA
J D Erickson, L Botto, National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, Georgia, USA
S E Vollset, University of Bergen, Bergen, Norway
P Mastroiacovo, International Centre on Birth Defects, Rome, Italy

Correspondence to: Dr L D Botto, National Center on Birth Defects and Developmental Disabilities, Mailstop F-45, Centers for Disease Control and Prevention, 4770 Buford Highway NE, Atlanta, GA 30341, USA; LBotto@cdc.gov

REFERENCES

1 **Kang SS**, Wong PW, Susmano A, Sora J, Norusis M, Ruggie N. Thermolabile methylenetetrahydrofolate reductase: an inherited risk factor for coronary artery disease. *Am J Hum Genet* 1991;**48**:536-45.
2 **Frosst P**, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, Boers GJ, den Heijer M, Kluijtmans LA, van den Heuvel LP, *et al*. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat Genet* 1995;**10**:111-13.
3 **Botto LD**, Yang Q. 5,10-Methylenetetrahydrofolate reductase gene variants and congenital anomalies: a HuGE review. *Am J Epidemiol* 2000;**151**:862-77.
4 **Ueland PM**, Hustad S, Schneede J, Refsum H, Vollset SE. Biological and clinical implications of the MTHFR C677T polymorphism. *Trends Pharmacol Sci* 2001;**22**:195-201.
5 **Ray JG**, Laskin CA. Folic acid and homocyst(e)ine metabolic defects and the risk of placental abruption, pre-eclampsia and spontaneous pregnancy loss: a systematic review. *Placenta* 1999;**20**:519-29.
6 **Ames BN**. Cancer prevention and diet: help from single nucleotide polymorphisms. *Proc Natl Acad Sci USA* 1999;**96**:12216-18.
7 **Ames BN**, Elson-Schwab I, Silver EA. High-dose vitamin therapy stimulates variant enzymes with decreased coenzyme binding affinity (increased K(m)): relevance to genetic disease and polymorphisms. *Am J Clin Nutr* 2002;**75**:616-58.
8 **Rosenblatt DS**. Methylenetetrahydrofolate reductase. *Clin Invest Med* 2001;**24**:56-9.
9 **Khoury MJ**, Little J. Human genome epidemiologic reviews: the beginning of something HuGE. *Am J Epidemiol* 2000;**151**:2-3.
10 **Rosenberg N**, Murata M, Ikeda Y, Opare-Sem O, Zivelin A, Geffen E, Seligsohn U. The frequent 5,10-methylenetetrahydrofolate reductase C677T polymorphism is associated with a common haplotype in Whites, Japanese, and Africans. *Am J Hum Genet* 2002;**70**:758-62.
11 **Vollset SE**. Confidence intervals for a binomial proportion. *Stat Med* 1993;**12**:809-24.
12 **Brown LD**, Cai TT, DanGupta A. Interval estimation for a binomial proportion. *Stat Sci* 2001;**16**:101-33.
13 **Mills JL**, Kirke PN, Molloy AM, Burke H, Conley MR, Lee YJ, Mayne PD, Weir DG, Scott JM. Methylenetetrahydrofolate reductase thermolabile variant and oral clefts. *Am J Med Genet* 1999;**86**:71-4.
14 **Shaw GM**, Rozen R, Finnell RH, Wasserman CR, Lammer EJ. Maternal vitamin use, genetic variation of infant methylenetetrahydrofolate reductase, and risk for spina bifida. *Am J Epidemiol* 1998;**148**:30-7.
15 **Rajkovic A**, Mahomed K, Rozen R, Malinow MR, King IB, Williams MA. Methylenetetrahydrofolate reductase 677 C>T polymorphism, plasma folate, vitamin B(12) concentrations, and risk of preeclampsia among black African women from Zimbabwe. *Mol Genet Metabol* 2000;**69**:33-9.
16 **Ubbink JB**, Christianson A, Bester MJ, Van Allen MI, Venter PA, Delport R, Blom HJ, van der Merwe A, Potgieter H, Vermaak WJ. Folate status, homocysteine metabolism, and methylene tetrahydrofolate reductase genotype in rural South African blacks with a history of pregnancy complicated by neural tube defects. *Metabolism* 1999;**48**:269-74.
17 **Wilcken DE**, Wang XL, Sim AS, McCredie RM. Distribution in healthy and coronary populations of the methylenetetrahydrofolate reductase (MTHFR) C677T mutation. *Arterioscler Thromb Vasc Biol* 1996;**16**:878-82.
18 **Infante-Rivard C**, Rivard GE, Yotov WV, Genin E, Guiguet M, Weinberg C, Gauthier R, Feoli-Fonseca JC. Absence of association of thrombophilia polymorphisms with intrauterine growth restriction. *N Engl J Med* 2002;**347**:19-25.
19 **Mutchinick OM**, Lopez MA, Luna L, Waxman J, Babinsky VE. High prevalence of the thermolabile methylenetetrahydrofolate reductase variant in Mexico: a country with a very high prevalence of neural tube defects. *Mol Genet Metabol* 1999;**68**:461-7.
20 **de Franchis R**, Buoninconti A, Mandato C, Pepe A, Sperandeo MP, Del Gado R, Capra V, Salvaggio E, Andria G, Mastroiacovo P. The C677T mutation of the 5,10-methylenetetrahydrofolate reductase gene is a moderate risk factor for spina bifida in Italy. *J Med Genet* 1998;**35**:1009-13.
21 **Weitkamp LR**, Tackels DC, Hunter AG, Holmes LB, Schwartz CE. Heterozygote advantage of the MTHFR gene in patients with neural-tube defect and their relatives. *Lancet* 1998;**351**:1554-5.
22 **International Clearinghouse for Birth Defects Monitoring Systems**. *Annual report 2001*. Rome: International Centre for Birth Defects, 2001.
23 **Moore CA**, Li S, Li Z, Hong SX, Gu HQ, Berry RJ, Mulinare J, Erickson JD. Elevated rates of severe neural tube defects in a high-prevalence area in northern China. *Am J Med Genet* 1997;**73**:113-18.
24 **Wald DS**, Law M, Morris JK. Homocysteine and cardiovascular disease: evidence on causality from a meta-analysis. *BMJ* 2002;**325**:1202.
25 **Mukherjee M**, Joshi S, Bagadi S, Dalvi M, Rao A, Shetty K. A low prevalence of the C677T mutation in the methylenetetrahydrofolate reductase gene in Asian Indians. *Clin Genet* 2002;**61**:155-9.
26 **Hobbs CA**, Sherman SL, Yi P, Hopkins SE, Torfs CP, Hine RJ, Pogribna M, Rozen R, James SJ. Polymorphisms in genes involved in folate metabolism as maternal risk factors for Down syndrome. *Am J Hum Genet* 2000;**67**:623-30.
27 **O'Leary VB**, Parle-McDermott A, Molloy AM, Kirke PN, Johnson Z, Conley M, Scott JM, Mills JL. MTRR and MTHFR polymorphism: link to Down syndrome? *Am J Med Genet* 2002;**107**:151-5.
28 **James SJ**, Pogribna M, Pogribny IP, Melnyk S, Hine RJ, Gibson JB, Yi P, Tafoya DL, Swenson DH, Wilson VL, Gaylor DW. Abnormal folate metabolism and mutation in the methylenetetrahydrofolate reductase gene may be maternal risk factors for Down syndrome. *Am J Clin Nutr* 1999;**70**:495-501.
29 **Hassold TJ**, Burrage LC, Chan ER, Judis LM, Schwartz S, James SJ, Jacobs PA, Thomas NS. Maternal folate polymorphisms and the etiology of human nondisjunction. *Am J Hum Genet* 2001;**69**:434-9.
30 **Rozen R**. Genetic predisposition to hyperhomocysteinemia: deficiency of methylenetetrahydrofolate reductase (MTHFR). *Thromb Haemost* 1997;**78**:523-6.